IN THE DRAWINGS:

The applicant has amended Fig. 1 to include the voice activity detector (VAD), as shown red in the replacement of Fig. 1. This is fully supported in the original application in particular in lines 7-8, page 4.

REMARKS

All the pending claims 1 – 4 are rejected by the Examiner under 35 U.S.C. §103(a) over Ogata, et al. (JP 06-062400), Kamata, et al. (U.S. Patent No. 5,953,050) in view of Zhou (US Patent No. 5,512,939), as well as under the judicially created doctrine of double patenting over U.S. Patent No. 5,914,747. The applicant has deleted claims 1 – 4 without prejudice and added claims 5 – 15 for further examination. The applicant believes that the added claims 5 – 15 have overcome the rejections of the Examiner, as explained below.

A brief explanation on the present invention is believed helpful in recognizing the differences between the invention and the disclosures in the cited patents. In a videoconference system, it is useful and sometimes important to determine which conferee is speaking so that he or she can be displayed visually differently to other conferees, and the bandwidth allotment may also be optimized accordingly. The present invention teaches a novel technique with which the videoconference system may precisely determine whether a conferee is speaking by analyzing the consistency between a visual lip movement of the conferee and the audio signal from the conference station in which the conferee is located. This distinguishing feature is now explicitly defined in all the added independent claims 5, 11, 14 and 15, and is supported in the specification at page 4, lines 20-22. In other words, the present invention makes use of not only both the lip movement and the audio signal, but also the correlation between the lip movement and the audio signal, to determine whether the conferee is speaking. This is thus much more precisely than

those techniques used in the cited patents in which only the audio signal, only the lip movement or a simple combination of both is used to determine the speaking conferee.

In particular, both Ogata and Kamata use only audio signal from a conference station to determine whether the conferee in the station is speaking. This cannot work properly in some circumstances where the audio signal in the station is not a speech. For example, the major audio signal at the station may be music or noise in the background, but not a speech from the conferee.

In Zhou, the lip movement of a conferee is also used to determine whether the conferee is speaking. In particular, when the lips of a conferee are moving, this conferee is determined to be talking. However, this is not precise because the movement of the lips does not always make a speech. For example, the conferee may be yawning, or may be eating something. Even a coexistence of the lip movement and the audio signal can not simply determine the speaking conferee correctly because the audio signal may come from some background sound such as music or noise. In fact, as described throughout the disclosure of Zhou, when the lips of a conferee are moving, the conferee can only be determined to be "most likely" talking. Nowhere in Zhou is there found any discussion of a consistency between the lip movement and the audio signal for determining the speaking conferee, nor does Zhou imply so. Therefore, the present invention cannot be concluded from Zhou or its combination with Ogata and Kamata.

Thus, the applicant believes that the added independent claims 5, 11, 14 and 15 are not obvious over Zhou, Ogata, Kamata and/or their combination, and are therefore patentable. At least for the same reasons, claims 6 – 10, which are dependent to claim 1, and claims 12- 13, which are dependent to claim 11, are also patentable.

The applicant also respectfully believes that new claims 5 – 15 do not create a question of double patenting over U.S. Patent No. 5,914,747. The claims of the '747 patent are directed to

an invention in which the video signal from a conference station is diminished when the conferee is detected to be absence from the proximity of the station, so as to save bandwidth. None of these claims include the features discussed above relevant to this application, that the conferee is determined to be "speaking" by analyzing the consistency between a visual lip movement of the conferee and the audio signal from the conference station in which the conferee is located. In fact, as explicitly defined in its claims, the '747 patent is not directed to whether the conferee is speaking, but whether the conferee is located proximate to the conference station.

Applicant believes the application as amended is now in good condition for allowance, and reconsideration is here respectfully requested in view of the amendments and the above remarks. The Examiner is authorized to deduct additional fees believed due from our Deposit Account No. 11-0223.

Respectfully submitted,

KAPLAN & GILMAN, L.L.P.
900 Route 9 North
Woodbridge, New Jersey 07095
Telephone (732) 634-7634

DATED: October 16, 2001

Michael R. Gilman
(Reg. No. 34,826)

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal service as first class mail, in a postage prepaid envelope, addressed to Box Non-Fee Amendment, Commissioner for Patents, Washington, D.C. 20231 on October 16, 2001.

Dated October 16, 2001

Signed Paula M. Halsey

Print Name Paula M. Halsey

Replacement Pages of Added Claims 5 - 15

5. A videoconferencing system comprising:

a conference bridge for interconnecting a plurality of remotely located videoconference stations;

means for determining whether a conferee is speaking by analyzing a consistency between a visual lip movement of said conferee and an audio signal from a conference station in which said conferee is located; and

means for visually altering an image of said conferee displayed in other conference stations if said conferee is determined to be speaking.

6. The videoconference system of claim 5 wherein said means for determining whether said conferee is speaking comprises a voice activity detector.

7. The videoconference system of claim 6 wherein said voice activity detector is implemented at each of said conference stations.

8. The videoconference system of claim 6 wherein said voice activity detector is implemented at said conference bridge.

9. The videoconference system of claim 6 wherein said voice activity detector includes image analysis and recognition software.

10. The videoconference system of claim 5 wherein said means for visually altering said image comprises means for highlighting a border around said image of said conferee determined to be speaking.

11. A videoconference station comprising:

a transmitter to transmit a combined audio video signal to a videoconference bridge; and

means for determining whether a conferee located at said videoconference station is speaking by analyzing a consistency between a visual lip movement of said conferee and an audio signal at said station.

12. The videoconference station of claim 11 wherein said means for determining whether said conferee is speaking is a voice activity detector.

13. The videoconference station of claim 12 wherein said voice activity detector includes image analysis and recognition software.

14. A method of displaying images of a plurality of conferees in a videoconference system, comprising:

determining whether a conferee is speaking by analyzing a consistency between a visual lip movement of said conferee and an audio signal from a conference station in which said conferee is located; and

visually altering an image of said conferee that is displayed to other conferees when said conferee is determined to be speaking.

15. A method of determining whether a conferee in a videoconference is speaking, comprising analyzing a consistency between a visual lip movement of said conferee and an audio signal from a conference station in which said conferee is located.

Approved
G.Le 4/19/02

FIG. 1